W114–W118 *Nucleic Acids Research, 2008, Vol. 36, Web Server issue* *Published online 12 May 2008*
doi:10.1093/nar/gkn297

pssRNAMiner: a plant short small RNA regulatory cascade analysis server

Xinbin Dai and Patrick Xuechun Zhao*

Plant Biology Division, The Samuel Robert Noble Foundation, Ardmore OK 73401, USA

Received February 2, 2008; Revised April 16, 2008; Accepted April 29, 2008

ABSTRACT

In plants, short RNAs including ~21-nt microRNA (miRNA) and 21-nt *trans*-acting siRNA (ta-siRNA) compose a 'miRNA → ta-siRNA → target gene' cascade pathway that regulates gene expression at the posttranscriptional level. In this cascade, biogenesis of ta-siRNA clusters requires 21-nt intervals (i.e. phasing) and miRNA (phase-initiator) cleavage sites on its TAS transcript. Here, we report a novel web server, pssRNAMiner, which is developed to identify both the clusters of phased small RNAs as well as the potential phase-initiator. To detect phased small RNA clusters, the pssRNAMiner maps input small RNAs against user-specified transcript/genomic sequences, and then identifies phased small RNA clusters by evaluating *P*-values of hypergeometric distribution. To identify potential phase-initiators, pssRNAMiner aligns input phase-initiators with transcripts of TAS candidates using the Smith–Waterman algorithm. Potential cleavage sites on TAS candidates are further identified from complementary regions by weighting the alignment expectation and its distance to detected phased small RNA clusters. The pssRNAMiner web server is freely available at http://bioinfo3.noble.org/pssRNAMiner/.

INTRODUCTION

In plants, small regulatory RNAs including miRNAs, heterochromatic siRNAs (hc-siRNAs), repeat-associated siRNAs (ra-siRNAs), natural sense–antisense transcript siRNAs and *trans*-acting siRNAs (ta-siRNAs) are involved in regulating gene expression through various mechanisms (1). Recent studies have demonstrated that these various small RNA molecules, in combination with cellular transcription factors, form the basis of network responsible for regulating the cellular gene expression.

The ta-siRNA, a newly identified class of 21-nt short siRNAs, play an essential role in bridging the miRNA and siRNA pathways, which were previously believed to be two independent processes (2–5). Phase-initiators direct the cleavage of ta-siRNA gene (TAS) primary transcripts (precursors) and subsequently initiate the production of ta-siRNA clusters (4). To date, only a few known phase-initiators are ta-siRNAs as most of reported phase-initiators are miRNAs. Following cleavage of the precursor, the 3′ (or 5′)-cleavage products are converted into double-stranded RNA (dsRNA) by RDR6 and SGS3. The dsRNA is then processed into 21-nt increments, relative to the original cleavage site on both strands, by DCL4 in order to produce ta-siRNA clusters (3,4). The ta-siRNAs form an essential component of the RISC complex responsible for guiding AGO-dependent cleavage of the target transcript. Based on these data, several researchers have proposed a 'miRNA → ta-siRNA → target gene' cleavage cascade as an element of gene regulatory network in the model plant *Arabidopsis* (4–7). In this paper, we describe the development of the pssRNAMiner, a plant short small RNA regulatory cascade analysis server that is able to identify potential ta-siRNAs clusters and their phase-initiators.

While miRNA biogenesis is dependent on the hairpin structure of the precursor, the formation of siRNA molecules occurs as a result of the processing of dsRNA by RNA-dependent RNA polymerases (RDRs). Therefore, as compared to miRNA, there is no effective computational approach to identify siRNAs from either genomic or transcript sequence. Since 21-nt phased ta-siRNAs are generated from TAS precursors, it is possible to map known small RNA molecules to candidate TAS precursors and then cluster these small RNAs based on their 21-nt phase properties (8,9). By evaluating the *P*-value of phase features, Chen *et al.* (6) identified and validated a number of known, and several previously unknown, ta-siRNA clusters as well as the associated TAS gene loci in *Arabidopsis*.

TAS precursors require at least one valid cleavage site, which is targeted by a phase-initiator to generate ta-siRNAs. While some TAS genes possess multiple regions that are complementary with the phase-initiator, only one region effectively guides cleavage *in vivo*. A 'two-hit trigger' mechanism was recently proposed to correlate the number of complementary regions with the activity of

*To whom correspondence should be addressed. Tel: +1 580 224 6725; Fax: +1 580 224 6692; Email: pzhao@noble.org

phased-initiators based on evidences from studies performed in moss and *Arabidopsis* (7). Therefore, it is anticipated that analysis of both complementary regions as well as valid cleavage sites in TAS candidate precursors would facilitate screening of phased small RNA clusters.

In this study, we describe the development of pssRNAMiner, a web-based server which identifies ta-siRNA clusters as well as their potential phase-initiators. This program requires that the user submit a set of small RNAs and specify one of listed transcript/genomic libraries for mapping. To identify phase-initiators, the user must submit at least one small RNA as candidate phase initiator. The pssRNAMiner is able to identify phased small RNA clusters as ta-siRNA candidates by evaluating the *P*-values of hypergeometric distribution. Furthermore, pssRNAMiner has the ability to identify potential phase-initiators based on the user input. To date, pssRNAMiner hosts 29 transcript/genomic sequence libraries from 20 species.

METHODS

Detection of phased small RNA clusters

The pssRNAMiner improves a previously described method of evaluating the *P*-value of random hypergeometric distribution to detect phased small RNA clusters (6). First, pssRNAMiner maps the input small RNAs on transcript sequences and records each position. Then, as described by Chen *et al*. (6), the algorithm slides on both strands of the transcript sequence to search each mapped small RNA and count the number of phased/nonphased positions with small RNA hits in a 231 bp region downstream of the 5′ start site of the small RNA. Equations (1) and (2), revised from Chen *et al*. (6), are used to calculate the *P*-value of phased small RNA clusters on the basis of a random hypergeometric distribution. Since the cleavage of phased small RNAs often occurs within 1–2 nt of the phased positions (8), we introduced a variable, *s*, in Equation (1) to reflect this shift. The addition of this variable will have the added effect of reducing the total number of nonphased position within the 231 bp region.

$$\Pr(X=k)=\frac{\binom{400-21\times 2\times s}{n-k}\binom{21}{k}}{\binom{461-21\times 2\times s}{n}} \quad \mathbf{1}$$

n: Number of total positions having small RNA hits in 231-bp region;
k: Number of phase positions having small RNA hits in 231-bp region;
s: Maximum allowed offset from phase position

$$P\text{-value}: p(k)=\sum_{X=k}^{\min(n,21)}\Pr(X) \quad \mathbf{2}$$

Analysis of cleavage sites guided by phase-initiator

In addition to detecting phased small RNA clusters, pssRNAMiner has the ability to predict whether the input candidate phase-initiators have the potential to guide the cleavage of TAS candidate precursors and trigger the biogenesis of phased small RNA clusters identified in the first step.

To perform this analysis, pssRNAMiner first searches the complementary regions between the phase-initiator and TAS candidate precursor using the Smith–Waterman algorithm (10). It then ranks the alignment of complementary regions based on a scoring scheme described by Zhang (11,12). Since the Smith–Waterman algorithm is only able to identify the optimal alignment for each pair of query/target sequences, an iterative algorithm has been developed to obtain as many complementary regions as possible. During each iteration, the identified complementary region of the precursor sequences are masked by '*N*' letters for the next iterative search until no further complementary region(s) can be identified in a given query/target sequence.

Next, the server identifies whether these complementary regions have valid cleavage sites, which can result in the production of phased small RNA clusters. We applied two conditions in this analysis. First, the user-defined cleavage site in the complementary region must be located within a region calculated by $[B_1-D,B_2+D]$, where B_1, start position of phased small RNA cluster on the TAS precursor; B_2, end position of phased small RNA cluster; *D*, a user-specific maximum distance between cleavage site and start site of small RNA cluster. Second, the distance between the cleavage site and start site of the small RNA cluster must meet the 21-nt phase property (i.e. multiples of 21 nt).

Architecture and implementation of pssRNAMiner

The pssRNAMiner consists of two independent components, a backend pipeline that is responsible for directing the core functions and a web server. The core functions, which include the calculation of the *P*-value and analysis of cleavage sites directed by the phase-initiator, are performed by the backend pipeline written in Java and PERL. To perform a search for complementary regions, a third party software *ssearch* has been used for Smith–Waterman alignment. The web server was developed on top of a SQLite database. Scripts used to generate web interfaces and search results were written in PHP.

USER INTERFACES

To identify 21-nt phased small RNA clusters as ta-siRNA candidates from the existing small RNA dataset, the user is required to submit small RNA sequences in either a simple sequence format or multi-FASTA format. The pssRNAMiner is able to recognize small RNA sequences between 17 and 28 nt in length but more readily identifies sequences of 21 nt in length. In addition, the user needs to specify a transcript/genomic library for mapping (Figure 1a). Following submission, pssRNAMiner maps each of the small RNAs onto the transcript/genomic sequence and then, calculates a *P*-value, in order to evaluate the potential phased small RNA cluster. The pssRNAMiner lists the candidate clusters having *P*-values

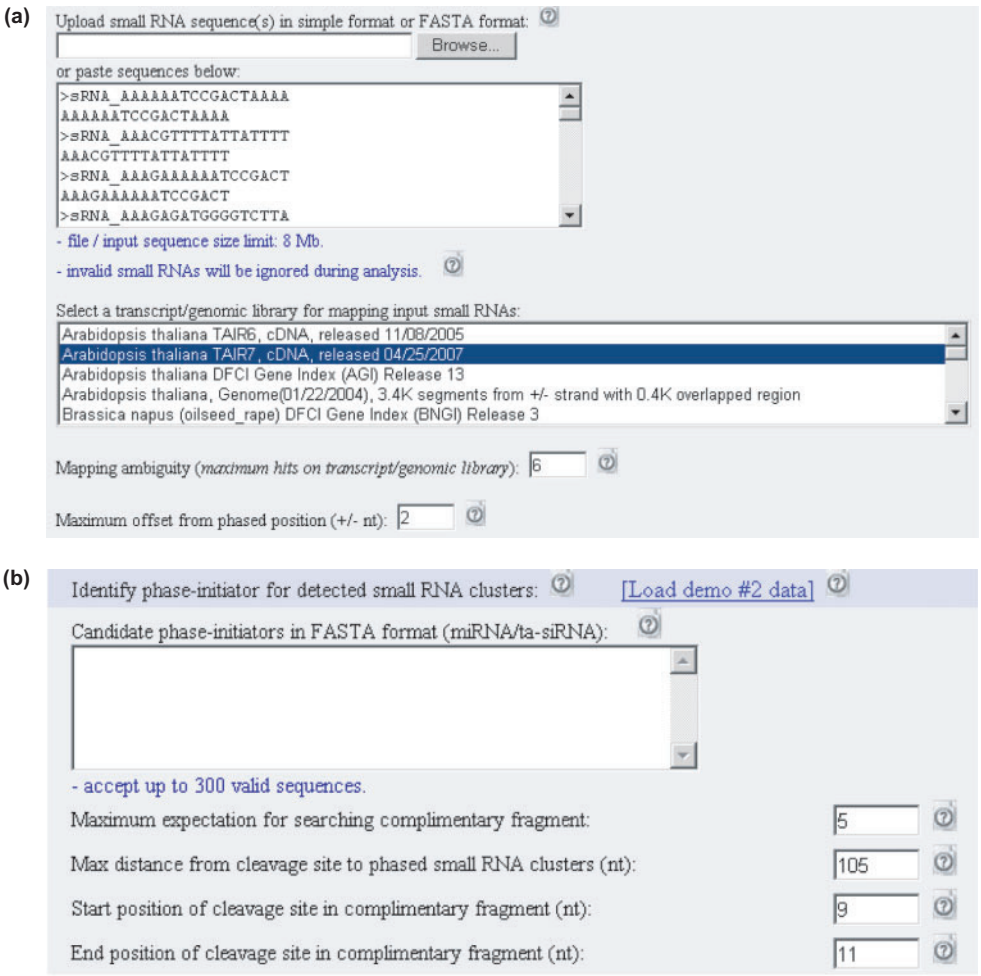

Figure 1. (**a**) Input interface for searching phased small RNA clusters. (**b**) Input interface for optional identification of phase-initiator for detected small RNA clusters.

lower than a user specified cutoff threshold (Figure 2a). The list view page enables the use of a filtering/searching function, in order to remove false positive clusters, such as transposon or ribosomal RNAs. In addition, this function enables the user to list only those clusters of interest by filtering based on specific annotation keywords or IDs. The users can further inspect the details of each cluster, including phased/nonphased small RNA sequences, the location of small RNAs on TAS candidate precursors, functional annotation of TAS candidate and valid cleavage site, etc. by clicking on the individual cluster ID for a detailed view (Figure 2b and c). The users are able to further check potential target sequences of these detected ta-siRNA candidates by clicking 'search' at the end of detailed view.

Based on user preferences, pssRNAMiner can also identify potential cleavage site guided by phase-initiator on TAS candidate precursor. For pssRNAMiner to perform this function, it is necessary for the user to submit potential phase-initiators in multi-FASTA format and specify the maximum expectation for screening the complementary region between the phase-initiators and TAS candidate precursors. In addition, it is necessary for the user to specify the expected cleavage range in the complementary region (generally between 9 and 11 nt on miRNA sequence) (Figure 1b). After submission, the server aligns input phase-initiators with TAS candidate precursors to locate complementary regions. Next, the server searches valid cleavage sites from the complementary regions based on the distance between the cleavage site and the phased small RNA clusters and output a web page that lists clusters with cleavage site information (Figure 2a–c).

To demonstrate the effectiveness of pssRNAMiner and to facilitate its use, we have integrated a number of published small RNA datasets into pssRNAMiner. These data can be found under the 'dataset' sub-menu. Each dataset has been pre-run against their corresponding transcript/genomic sequences, and therefore users may click links on the right-hand column of dataset table to view these precalculated results.

PERFORMANCE

Finally, to evaluate the performance of pssRNAMiner, we downloaded RDR small RNA MPSS libraries from the

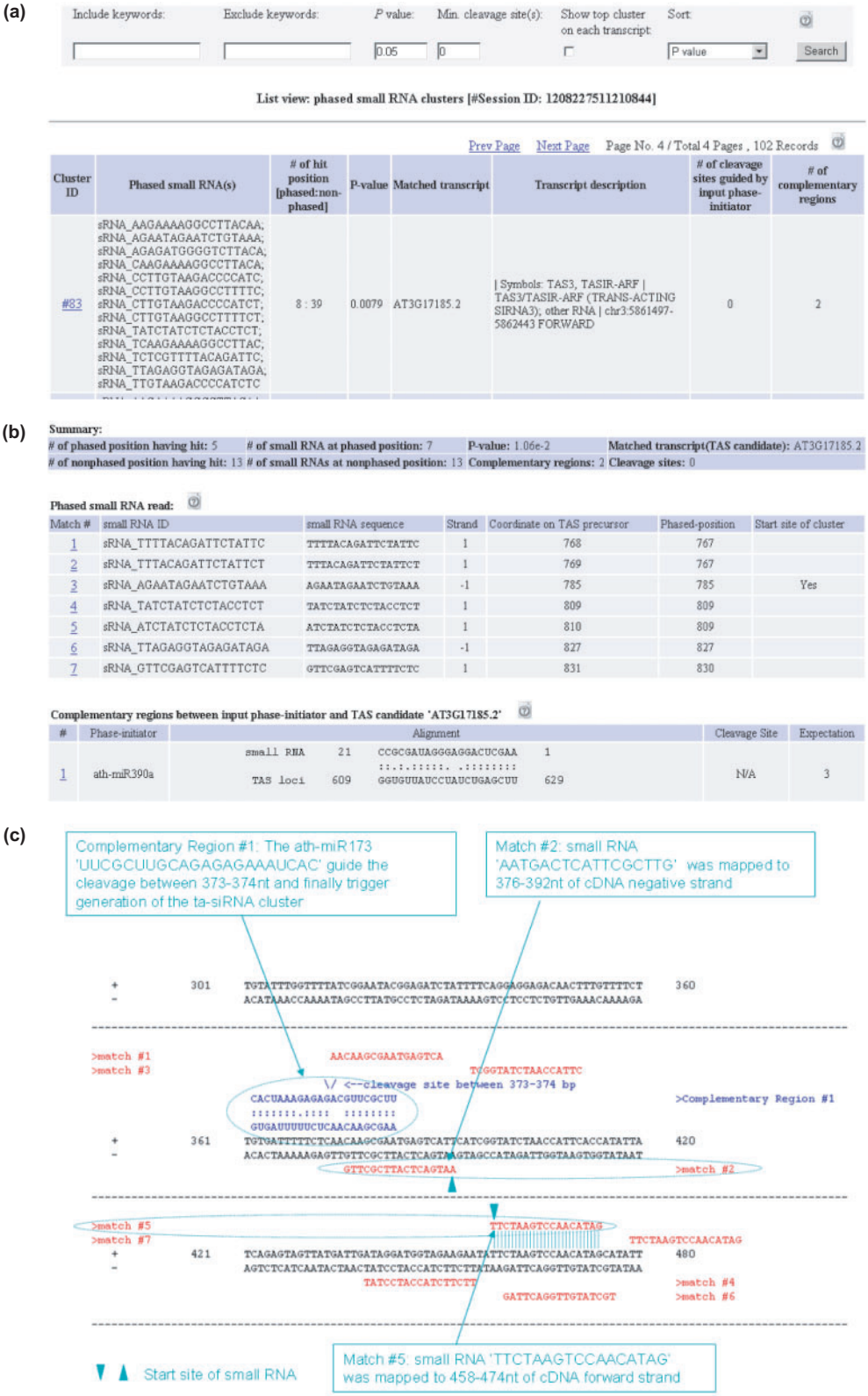

Figure 2. (**a**) Output: list view of phased small RNA clusters with phase-initiator cleavage sites on TAS candidate. (**b**) Output: detailed view of a specific phased small RNA clusters with phase-initiator cleavage sites on a TAS candidate. (**c**) Output: detailed view of the alignment of specific phased small RNA clusters with phase-initiator cleavage sites on a TAS candidate.

Arabidopsis MPSS Plus database (http://mpss.udel.edu/at/) (13,14). A total of 11 767 distinct small RNA signatures were analyzed based on the *Arabidopsis* TAIR7 cDNA release (ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR7_genome_release/). When these data were investigated, pssRNAMiner detected 124 significantly phased small RNA clusters ($P < 0.005$) at 19 gene loci. Moreover, pssRNAMiner detected all the reported ta-siRNA gene loci with significant *P*-value, including TAS1a (AT2G27400), TAS1b (AT1G50055), TAS1c (AT2G39675), TAS2 (AT2G39681), TAS3a (AT3G17185) and PPR proteins (AT1G63080 and AT1G63130) (see Supplementary Material for the list of detected phased RNA clusters and gene loci). To investigate the ability of pssRNAMiner to identify valid cleavage site of phase-initiators, a reported phase-initiator, miRNA 'ath-miR390', was submitted for cleavage site analysis on TAS candidate precursors. The pssRNAMiner reported two complementary regions (expectation ≤ 5) with one region having a valid cleavage site on TAS3a. These results are consistent with previously published studies that demonstrated that the ath-miR390 initiates production of phased small RNA clusters on TAS3a by guiding precursor cleavage (2,4) (see detailed alignment in the Supplementary Material).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to the members of our group, colleagues and our external beta testers, especially Vikram Agarwal, who have rigorously tested the pssRNAMiner and provided valuable feedback. Financial support for this project and funding to pay the Open Access publication charges for this article was provided by The Samuel Roberts Noble Foundation.

Conflict of interest statement. None declared.

REFERENCES

1. Borsani,O., Zhu,J., Verslues,P.E., Sunkar,R. and Zhu,J.-K. (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell*, **123**, 1279–1291.
2. Vazquez,F., Vaucheret,H., Rajagopalan,R., Lepers,C., Gasciolli,V., Mallory,A.C., Hilbert,J.-L., Bartel,D.P. and Crete,P. (2004) Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. *Mol. Cell*, **16**, 69–79.
3. Peragine,A., Yoshikawa,M., Wu,G., Albrecht,H.L. and Poethig,R.S. (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. *Genes Dev.*, **18**, 2368–2379.
4. Allen,E., Xie,Z., Gustafson,A.M. and Carrington,J.C. (2005) microRNA-Directed phasing during trans-acting siRNA biogenesis in plants. *Cell*, **121**, 207–221.
5. Yoshikawa,M., Peragine,A., Park,M.Y. and Poethig,R.S. (2005) A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. *Genes Dev.*, **19**, 2164–2175.
6. Chen,H.-M., Li,Y.-H. and Wu,S.-H. (2007) Bioinformatic prediction and experimental validation of a microRNA-directed tandem trans-acting siRNA cascade in Arabidopsis 10.1073/pnas.0611119104. *Proc. Natl Acad. Sci.*, **104**, 3318–3323.
7. Axtell,M.J., Jan,C., Rajagopalan,R. and Bartel,D.P. (2006) A two-hit trigger for siRNA biogenesis in plants. *Cell*, **127**, 565–577.
8. Rajagopalan,R., Vaucheret,H., Trejo,J. and Bartel,D.P. (2006) A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes Dev.*, **20**, 3407–3425.
9. Howell,M.D., Fahlgren,N., Chapman,E.J., Cumbie,J.S., Sullivan,C.M., Givan,S.A., Kasschau,K.D. and Carrington,J.C. (2007) Genome-wide analysis of the RNA-dependent RNA polymerase6/dicer-like4 pathway in Arabidopsis reveals dependency on miRNA- and tasiRNA-directed targeting. *Plant Cell*, **19**, 926–942.
10. Smith,T. and Waterman,M. (1981) Identification of common molecular subsequences. *J. Mol. Biol.*, **147**, 195–197.
11. Zhang,Y. (2005) miRU: an automated plant miRNA target prediction server. *Nucleic Acids Res.*, **33**, W701–704.
12. Brennecke,J., Stark,A., Russell,R.B. and Cohen,S.M. (2005) Principles of microRNA target recognition. *PLoS Biol.*, **3**, e85.
13. Meyers,B.C., Tej,S.S., Vu,T.H., Haudenschild,C.D., Agrawal,V., Edberg,S.B., Ghazal,H. and Decola,S. (2004) The use of MPSS for whole-genome transcriptional analysis in Arabidopsis. *Genome Res.*, **14**, 1641–1653.
14. Meyers,B.C., Galbraith,D.W., Nelson,T. and Agrawal,V. (2004) Methods for transcriptional profiling in plants. Be fruitful and replicate. *Plant Physiol.*, **135**, 637–652.